REVIEW 10.1111/j.1469-0691.2012.03781.x

Genomics of epidemic pathogens

K. Georgiades

Unité de Recherche en Maladies Infectieuses Tropicales Emergentes (URMITE), CNRS-IRD UMR 6236, Faculté de Médecine, Université de la Méditerranée, Marseille, France

Abstract

Virulence factors are thought to be responsible for the virulence capacity of pathogenic bacteria. However, epidemic bacteria were recently found to contain significantly fewer 'virulence factors' than non-epidemic species, and some of the most dangerous epidemic bacteria, such as *Mycobacteria* spp. and *Rickettsia* spp., have reduced genomes, and contain hundreds of degraded genes. Epidemic bacteria are actually highly specialized species, characterized by allopatric speciation, that, after adapting to their hosts, attempt to maintain a balance between gene gain and gene loss that favours gene loss, finally leading to genome reduction. Recent comparative genomic studies have demonstrated that the specialization of bacteria to eukaryotic cells is associated with massive gene loss. Furthermore, the 12 deadliest epidemic species for humankind have significantly smaller genomes, with fewer open reading frames, than less dangerous species. Epidemic species mostly lose genes related to metabolic activity, the production of energy, cell motility, and transcription. Epidemic bacteria also possess a damaged recombination and repair system and significantly more toxins than closely related non-pathogenic or non-epidemic species, and more toxin–antitoxin modules. Epidemic bacteria are therefore highly specialized species that are adapted to their hosts and characterized by extensive genome reduction. Except for toxins and toxin–antitoxin modules, which have a direct and measurable effect, other 'virulence factors' are factors associated with fitness in experimental models. Epidemic species are defined by a virulent genomic repertoire including both present and absent genes.

Keywords: Epidemic pathogens, gene loss, genomics, toxins, virulence
Article published online: 25 January 2012
Clin Microbiol Infect 2012; **18:** 213–217

Corresponding author: K. Georgiades, Université de la Méditerranée, Faculté de Médecine La Timone, 27 Bd Jean Moulin, 13385, Marseille Cedex 5, France
E-mail: popigeorgiades@hotmail.com

Introduction

Virulence factors are thought to be responsible for the virulence capacity of pathogenic bacteria [1]. This widespread and anthropocentric idea comes from the fact that deletion of genes in pathogens has a detrimental effect on their fitness and on their capability to cause disease [2]. Therefore, these removed genes were referred to as 'virulence factors' [3]. However, recent comparative studies have demonstrated that some of the most dangerous epidemic bacteria, such as *Mycobacterium leprae* and *Rickettsia prowazekii*, whose host range is extremely restricted, have reduced genomes and contain hundreds of degraded genes [4–10]. Furthermore, genes defined as coding for 'virulence factors' are also found in non-pathogenic species [11], and epidemic bacteria were found to contain significantly fewer 'virulence factors' than non-epidemic species [12]. In a study on a group of *Escherichia coli* strains (B2) involved in commensalism and in intestinal and extraintestinal pathogenesis, genes considered to be associated with virulence were found to be implicated in complex host–commensal niche colonization [13], providing evidence that virulence is a coincidental byproduct of commensalism [14].

If 'virulence factors' do not constitute the main characteristic of highly pathogenic bacteria, then what does? In this review, I examine recent genomic studies in an effort to determine the features that characterize the pathogenic capacity of the most dangerous epidemic bacteria.

Speciation

The most dangerous epidemic bacteria are specialized for eukaryotic cells. The notions of allopatry and sympatry are now used to describe the lifestyle of bacteria [15,16]. Sympatric bacteria live in the same environment, or geographical area, and therefore they exchange genes easily because they interact with many other bacteria of different phyla. For example, amoebae and arthropods constitute sympatric environments in which bacteria evolve [15,17]. These bacteria constitute species complexes and not bona fide species, and they often have the largest genomes, have more genes, are more resistant to physicochemical agents, and have better metabolic capacities [11]; they are therefore better adapted to their environment [18]. New metabolic abilities are gained by these complexes through horizontal gene transfer (HGT) [19]. Other events in microbial evolution, such as recombination, point mutations, and genome rearrangements, allow better adaptation to new niches without implying genetic isolation boundaries. In a general way, sympatric species create new genomic repertoires that form the reservoir of future specialists. Specialization results in allopatry, or geographical isolation, and a lack of gene acquisition through HGT. Allopatric speciation will therefore be characterized by an irreversible genomic size restriction. The DNA recombination and repair system is defective, and the introduction of deleterious mutations will lead to further loss of functions of genes. Deregulation will eventually lead to uncontrolled multiplication, and pathogenicity is demonstrated by destruction of the ecosystem. Most dangerous pathogenic epidemic bacteria are actually highly specialized species that, after adapting to their hosts, attempt to maintain a balance between gene gain and gene loss that favours gene loss, finally leading to genome reduction [12].

Gene Loss

Recent comparative genomic studies have demonstrated that the specialization of bacteria to eukaryotic cells is associated with massive gene loss [10,20] and the loss of 'virulence factors' [21]. Speciation is accompanied by a reduced number of ribosomal operons [22] and by the loss of 100 genes by all obligate intracellular bacteria, suggesting an irreversible association with the host [10]. In a study on the evolution of the *Rickettsiales*, it was demonstrated that evolution from a free-living lifestyle to an obligate intracellular one was associated with the loss of 2135 genes [23]. Furthermore, the 12 deadliest epidemic species for humankind have significantly smaller genomes, with fewer open reading frames, than less dangerous species [12].

One of the best examples of genomic reduction of epidemic bacteria is *R. prowazekii*, the agent of epidemic typhus. No virulence genes have been identified in its genome, and 24% of its small genome is composed of pseudogenes and non-coding DNA [24,25]. Intracellular motility, which has been considered to be a virulence factor for *Listeria monocytogenes* [26] and *Shigella* [15], is not found in *R. prowazekii*, which is completely immobile in the cytoplasm [27,28]. Genes coding for amino acid biosynthesis are lost from *R. prowazekii*, as are translation regulation factors; translation capacities are decreased [24]. In another study comparing *Rickettsia africae* with *Rickettsia rickettsii*, it was demonstrated that the loss of essential genes was a key factor involved in the development of pathogenicity [9].

Other examples of excessive gene loss in epidemic pathogens are provided by *Mycobacteria* spp. and, especially, *M. leprae*, which has the largest proportion of non-coding DNA [5]; only 49.5% of its genome contains protein-coding genes. The leprosy bacillus has lost about 2000 genes, including genes involved in biosynthetic pathways [29]. Studies on *Mycobacterium tuberculosis* have shown that deletion of genes confers a hypervirulent phenotype [30], and *Mycobacterium ulcerans* has become specialized with the loss of 'virulence factors' and immunogenes [31].

Another paradigm of gene loss occurred during the evolution of two other host-restricted species, *Bordetella pertussis* and *Bordetella parapertussis*. Metabolic pathways and regulatory networks were modified, resulting in virulence characteristics and effective host infection [32].

Finally, an outstanding example of a dangerous epidemic pathogen is *Shigella dysenteriae*. It is a clone from the *E. coli* complex that differs from *E. coli* in its poor phenotypic traits (extracellular immobility and inability to ferment lactose) [33,34]. The most plausible scenario is that *Shigella* evolved from the *E. coli* complex through a plasmid containing critical genes. Then, massive gene deletions followed that increased its virulence, and virulence genes were lost, as in other pathogenic bacteria [35]. A recent comparative genomic study demonstrated that generally, ten of 23 functional COG categories contain significantly fewer genes in the deadliest epidemic bacteria. These categories contain genes mostly related to metabolic activity, the production of energy, cell motility, and transcription. The most dangerous epidemic bacteria also possess a damaged recombination and repair system, and present an accumulation of poly(A) tails. These tails lead to an accumulation of stop codons and to gene degradation. Because the repair machinery of epidemic bacteria is deficient, polymerase errors will not be corrected,

and genes will be inactivated, eventually resulting in pseudogenization and total gene loss [12]. Epidemic pathogens are therefore deregulated and evidently not better armed than other bacteria.

In summary, a pathogenic species is not characterized by any virulence factors, except for toxins and toxin–antitoxin (TA) modules, as we will see below, but by extensive genome reduction resulting from extreme specialization and adaptation in a stable environment. In the pathogenic gene repertoire of most dangerous pathogenic bacteria, absent genes are as important as those that are present.

Toxins

Toxins are macromolecular substances that, when produced during infection, or when introduced into an organism, cause an impairment of physiological functions that leads to disease or even to the death of the infected organism [36]. Since 1888, toxins have been considered to be the ultimate virulence factors. They are classified in two categories: bacterial protein toxins, or exotoxins, and toxic lipopolysaccharide (LPS) complexes, or endotoxins. Exotoxins are secreted by living bacteria, and their production is usually specific to a particular bacterial species; the disease is associated with the toxin. The major symptoms associated with diseases caused by *Corynebacterium diphtheriae*, *B. pertussis*, *Vibrio cholerae*, *Bacillus anthracis* and *Clostridium botulinicum* are related to the activities of the toxins produced by these organisms [36]. Protein toxins have a chemical, direct and quantifiable action that establishes an aggressive strategy during microbial pathogenesis. In contrast, LPS complexes have non-specific actions that subvert the host's immune response, and they are liberated in the cytoplasm in the case of, for example, bacterial lysis [37]. In experimental models, LPS complexes act in a way that does not reflect bacterial virulence [38]. This is the reason why the virulence of species such as *Salmonella* and *Yersinia*, which do not liberate protein toxins, could not be explained.

In a comparative genomic study on the 14 deadliest epidemic bacteria of all times for humankind, all of the epidemic pathogens were found to have significantly smaller genomes than closely related non-pathogenic or non-epidemic species, clearly demonstrating extensive gene loss in these species; the only features found in significantly larger numbers in epidemic species were TA modules and toxins. Moreover, endotoxins appear to be more related to the pathogenic capacity of epidemic bacteria than exotoxins [39]. Therefore, the only proteins that could be considered to be playing a role as virulence factors are toxins.

TA Modules

TA modules were initially identified as plasmid stabilization factors, and it has also been proposed that they play a role in the control of protein expression [40–42]. These systems are described as addiction molecules; the toxin and antitoxin genes are found next to each other on the same operon, so any effort to eliminate one of the two will lead to the death of the bacterium. Therefore, these genes are not essential genes, but genes from which organisms simply cannot be separated [43]. In our comparative genomic study on the 14 deadliest bacterial species, we found that epidemic species contained significantly more TA modules than non-epidemic species [12,39]. These modules are found in epidemic species probably as a result of HGT from other bacteria that occurred before specialization of the epidemic species [39]. Previous studies also reported a high number of TA modules in epidemic species, such as *Yersinia pestis* [44], but their role in pathogenicity was not considered. Indeed, their role in the virulence capacity of bacteria is not clearly established yet; however, in a recent study in our laboratory, it was demonstrated that liberation of the toxin into the cytoplasm of infected cells can cause death of the cells by apoptosis [45], and after attempts to limit their translation, pathogenicity was initiated.

As in the case of endotoxins, TA modules have not previously been considered to be essential factors in the pathogenic capacity of epidemic bacteria, because they do not have a direct effect, like protein toxins. Recent evidence, however, allows us to consider them as pathogens' hidden defence weapons. We could therefore speculate that, in a general way, epidemic bacteria are not armed to kill; they do not have supplementary virulence factors, except for selfish elements that express toxicity when the bacterium is threatened.

Discussion

Bacterial species constitute melting pots from which specialized species arise regularly [46–48]. Non-specialized species constitute 'pre-species' found in a community that allows them to exchange genes. At some point, probably because of an ecological change, they will be specialized to a specific niche, and gene exchange will decrease. This specialization will lead to gene loss and deregulation. Most dangerous pathogenic bacteria are therefore highly specialized species, adapted to their hosts and characterized by extensive genome reduction [12]. It is becoming increasingly evident that the pathogenic capacity of bacteria is not the result of 'virulence factors' (Fig. 1). Except for TA modules and toxins,

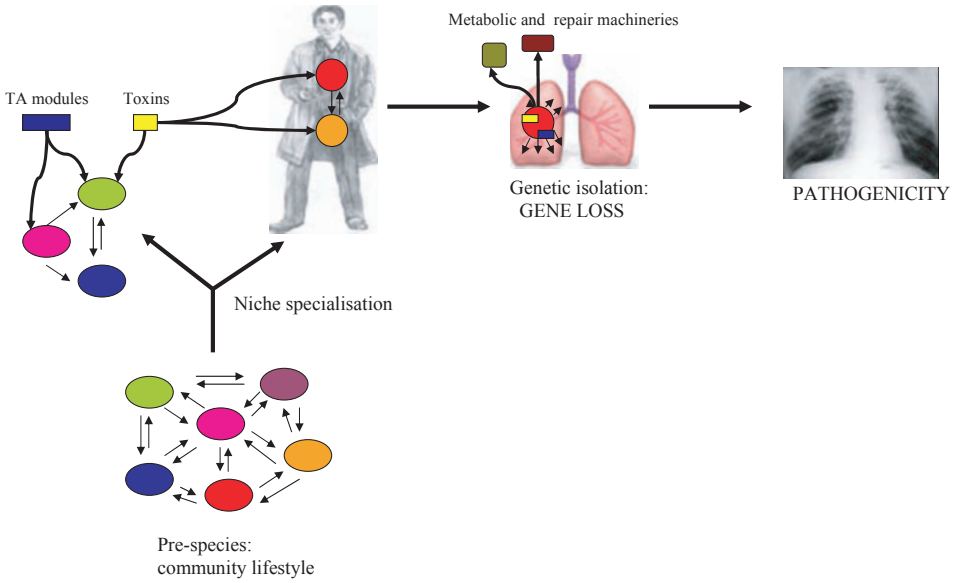

FIG. 1. Creation of an epidemic pathogen. Non-specialized bacterial species constitute 'pre-species' that have a community lifestyle and are able to exchange genes. After specialization, the number of horizontal gene transfers (HGTs) is reduced, until total genetic isolation has been acheived. Only gene loss is then possible, the bacterium is deregulated, and virulence capacity is initiated. The arrows show HGT and gene loss. TA, toxin–antitoxin.

TABLE 1. Genomic characteristics of epidemic bacteria and non-epidemic related species

Bacteria	Genome size (bp)	Toxins	TA modules
Mycobacterium leprae	3 268 203	1	0
Mycobacterium avium 104	5 475 491	0	1
Mycobacterium tuberculosis	4 411 532	4	38
Mycobacterium smegmatis	6 988 209	0	4
Rickettsia prowazekii	1 111 523	0	0
Rickettsia africae	1 290 917	0	10
Corynebacterium diphtheriae	2 488 635	2	0
Corynebacterium glutamicum	3 314 179	0	1
Treponema pallidum	1 138 012	0	0
Treponema denticola	2 843 201	0	1
Yersinia pestis orientalis	4 600 755	14	5
Yersinia pseudotuberculosis	4 744 671	6	0
Bordetella pertussis	4 086 186	33	0
Bordetella bronchiseptica	5 339 179	0	0
Streptococcus pneumoniae	2 160 837	2	4
Streptococcus agalacticae	2 160 267	1	2
Streptococcus pyogenes	1 852 442	21	2
Streptococcus suis	2 096 309	2	0
Salmonela Typhi	4 809 038	16	6
Salmonella Schwarzegrund	4 709 075	0	1
Shigella dysenteriae	4 369 232	2	2
Escherichia coli HS	4 643 537	0	1
Vibrio cholerae	4 033 460	11	13
Vibrio parahaemolyticus	5 165 770	0	4
Staphylococcus aureus	2 839 469	34	3
Staphylococcus haemolyticus	2 697 015	0	3
Neisseria meningitidis	2 272 360	7	6
Neisseria cinerea	1 872 373	3	2

TA, toxin–antitoxin.

which have a direct and measurable effect, other 'virulence factors' are factors associated with fitness in experimental models. Epidemic bacteria have much smaller genomes than other bacteria, they lack metabolic activity, they are completely deregulated, with deficient repair machineries, and they possess toxins, alone or coupled as TA modules [39] (Table 1). We conclude that epidemic species are defined by a virulent genomic repertoire including both present and absent genes. The capacity of epidemic species to obtain new characteristics is limited because of their genetic isolation. Therefore, any significant change in their ecosystem may result in the disappearance of the bacterium [49]. This is why we believe that current epidemic pathogens will probably disappear, but that they will be replaced by other bacteria already in contact with us, emerging from human commensals, animals, and the environment [50]. Indeed, the outbreak of bloody diarrhoea and the haemolytic–uraemic syndrome caused by an *E. coli* O104:H4 strain in Germany in May and June 2011 [51] illustrates the capacity of bacterial species to produce new combinations of genes, leading to the emergence of highly aggressive strains.

Transparency Declaration

The authors declare that there are no competing interests.

References

1. Wu HJ, Wang HJ, Jennings MP. Discovery of virulence factors of pathogenic bacteria. *Curr Opin Chem Biol* 2008; 12: 1–9.

2. ten Bokum AM, Movahedzadeh F, Frita R, Bancroft JG, Stoker GN. The case of hypervirulence through gene deletion in *Mycobacterium tuberculosis*. *Trends Microbiol* 2008; 16: 436–441.
3. Dobrindt U, Hochhut B, Hentschel U, Hackel J. Genomic islands in pathogenic and environmental microorganisms. *Nat Rev Microbiol* 2007; 7: 50–60.
4. Wixon J. Reductive evolution in bacteria: *Buchnera* sp., *Rickettsia prowazekii*, *Mycobacterium leprae*. *Comp Funct Genomics* 2001; 2: 44–48.
5. Cole ST, Eiglmeier K, Parkhill J *et al.* Massive gene decay in the leprosy bacillus. *Nature* 2001; 409: 1007–1011.
6. Sakharkar RK, Dhar KP, Chow TKV. Genome reduction in prokaryotic obligatory parasites of humans: a comparative analysis. *Int J Syst Evol Microbiol* 2004; 54: 1937–1941.
7. Andersson JO, Andersson SGE. Genome degradation is an ongoing process in *Rickettsia*. *Mol Biol Evol* 1999; 16: 1178–1191.
8. Ogata H, Audic S, Renesto-Audiffren P *et al.* Mechanisms of evolution in *Rickettsia conorii* and *Rickettsia prowazekii*. *Science* 2001; 293: 2093–2098.
9. Fournier PE, El Karkouri K, Leroy Q *et al.* Analysis of the *Rickettsia africae* genome reveals that virulence acquisition in *Rickettsia* species may be explained by genome reduction. *BMC Genomics* 2009; 10: 166–181.
10. Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. *Biol Direct* 2009; 4: 13.
11. Audic S, Robert C, Campagna B *et al.* Genome analysis of *Minibacterium massiliensis* highlights the convergent evolution of water living bacteria. *Plos Genet* 2007; 3: 1454–1463.
12. Georgiades K, Raoult D. Genomes of the most dangerous epidemic bacteria have a virulent repertoire characterized by fewer genes but more toxin–antitoxin modules. *PLoS ONE* 2011; 6: e17962.
13. Le Gall T, Clermont O, Gouriou S *et al.* Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic group *Escherichia coli* strains. *Mol Biol Evol* 2007; 24: 2373–2384.
14. Levin BR. The evolution and maintenance of virulence in microparasites. *Emerg Infect Dis* 1996; 2: 93–102.
15. Moliner C, Fournier PE, Raoult D. Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. *FEMS Microbiol Rev* 2010; 34: 281–294.
16. Raoult D, Boyer M. Amoebae as genitors and reservoirs of giant viruses. *Intervirology* 2010; 53: 321–329.
17. Thomas V, Greub G. Amoeba/amoebal symbiont genetic transfers: lessons from giant virus neighbours. *Intervirology* 2010; 53: 254–267.
18. Marco D. Metagenomics and the niche concepts. *Theory Biosci* 2008; 127: 241–247.
19. Lawrence J. Gene transfer in bacteria: speciation without species? *Theor Popul Biol* 2002; 61: 449–460.
20. Nierman WC, DeShazer D, Kim HS *et al.* Structural flexibility in the *Burkholderia mallei* genome. *Proc Natl Acad Sci USA* 2004; 101: 14246–14251.
21. Lescot M, Audic S, Robert C *et al.* The genome of *Borrelia recurrentis*, the agent of deadly louse-borne relapsing fever, is a degraded subset of tickborn *Borrelia duttonii*. *PLoS Genet* 2008; 4: e1000185.
22. Klappenbach JA, Dunbar JM, Schmidt TM. rRNA operon copy number reflects strategies of bacteria. *Appl Environ Microbiol* 2000; 66: 1328–1333.
23. Georgiades K, Merhej V, Pontarotti P, Raoult D. Gene gain and loss events in *Rickettsia* and *Orientia* species. *Biol Direct* 2011; 6: 6.
24. Andersson SG, Zomorodipour A, Andersson JO *et al.* The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 1998; 396: 133–140.
25. Bechah Y, Karkouri EK, Mediannikov O *et al.* Genomic, proteomic and transcriptomic analysis of virulent and avirulent *Rickettsia prowazekii* reveals its adaptive mutation capabilities. *Genome Res* 2010; 20: 655–663.
26. Goldberg MB, Theriot JA. *Shigella flexneri* surface protein Icsa is sufficient to direct actin-based motility. *Proc Natl Acad Sci USA* 1995; 92: 6572–6576.
27. Pollard TD. The cytoskeleton cellular motility and the reductionist agenda. *Nature* 2003; 422: 741–745.
28. Kleba B, Clark TR, Lutter EL, Ellison DW, Hackstadt T. Disruption of the *Rickettsia rickettsii* Sca2 autotransporter inhibits actin based motility. *Infect Immun* 2010; 78: 2240–2247.
29. Ribeiro-Guimaraes ML, Pessolani MCV. Comparative genomics of mycobacterial proteases. *Microb Pathog* 2007; 43: 173–178.
30. Casali N. Hypervirulent *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* 2003; 26: 15918–15923.
31. Demangel C, Stinear TP, Cole ST. Buruli ulcer: reductive evolution enhances pathogenicity of *Mycobacterium ulcerans*. *Nat Rev Microbiol* 2009; 7: 50–60.
32. Parkhill J, Sebaihia M, Preston A *et al.* Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat Genet* 2003; 35: 32–40.
33. Karaolis DKR, Lan RT, Reeves PR. Sequence variation in *Shigella sonnei* (*Sonnei*), a pathogenic clone of *Escherichia coli*, over 4 continents and 41 years. *J Clin Microbiol* 1994; 32: 796–802.
34. Pupo GM, Lan RT, Reeves PR. Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. *Proc Natl Acad Sci USA* 2000; 97: 10567–10572.
35. Maurelli AT, Fernandez RE, Bloch CA, Rode CK, Fasano A. Black holes and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*. *Proc Natl Acad Sci USA* 1998; 95: 3943–3948.
36. Alouf EJ. Bacterial toxins: methods and protocols. *Methods Mol Biol* 2000; 145: 1–26.
37. Merrell DS, Falkow S. Frontal and stealth attack strategies in microbial pathogenesis. *Nature* 2004; 430: 250–256.
38. Wesselink PR, Thodenvan Velzen SK, Makkes PC. Release of endotoxins in an experimental model simulating the dental root canal. *Oral Surg Oral Med Oral Pathol* 1978; 45: 789–795.
39. Georgiades K, Raoult D. Comparative genomics evidence that only protein toxins are tagging bad bugs. *Front Cell Infect Microbiol* 2011; 1: 7.
40. Szekeres S, Dauti M, Wilde C, Mazel D, Rowe-Magnus DA. Chromosomal toxin–antitoxin loci can diminish large scale genome reductions in the absence of selection. *Mol Microbiol* 2007; 63: 1588–1605.
41. Buts L, Lah J, Dao-Thi MH, Wyns L, Loris R. Toxin–antitoxin modules as bacterial metabolic stress managers. *Trends Biochem Sci* 2005; 30: 672–679.
42. Gerdes K, Christensen SK, Lobner-Olsen A. Prokaryotic TA stress response loci. *Nat Rev Microbiol* 2005; 3: 371–382.
43. D'Elia MA, Pereira MA, Brown DE. Are essential genes really essential? *Trends Microbiol* 2009; 17: 433–438.
44. Goulard C, Langrand S, Carniel E, Chauvaux S. The *Yersinia pestis* chromosome encodes active addiction toxins. *J Bacteriol* 2010; 192: 3669–3677.
45. Audoly G, Vincentelli R, Edouard S *et al.* Effect of rickettsial toxin VapC on its eukaryotic host. *PLoS ONE* 2011; 6: e26528.
46. Cohan FM. Sexual isolation and speciation in bacteria. *Genetica* 2009; 116: 359–370.
47. Doolittle WF, Papke RT. Genomics and the bacterial species problem. *Genome Biol* 2006; 7: 116.
48. Feil EJ. Linkage, selection and the clonal complex. In: Robinson DA, Falush D, Feil EJ, eds. *Bacterial population genetics in infectious disease*. Hoboken, NJ: John Wiley and Sons, 2010. doi: 10.1002/9780470600122.ch2.
49. Wolfe ND, Dunavan CP, Diamond J. Origins of major human infectious diseases. *Nature* 2007; 447: 279–283.
50. Georgiades K, Raoult D. Defining pathogenic bacterial species in the genomic era. *Front Microbiol* 2011; 1: 151.
51. Denamur E. The 2011 Shiga toxin-producing *Escherichia coli* O104:H4 German outbreak: a lesson of genomic plasticity. *Clin Microbiol Infect* 2011; 17: 1124–1125.